

## 119 years of the *Financial Times* archives on line

BASES No. 268, February 2010 - François Libmann

---

Following the announcement at Online Information in London, the archive of the *Financial Times* is now available online. Covering the period 1888 (when founded) to 2006 and gaining an additional year each year.

Supplied by Gale, a part of Cengage Learning, the database is brought to life using OCR (Optical Character Recognition) of the scanned editions (once they have been ironed, the most effective method of smoothing out the ancient editions!)

The total comprises approximately 790,000 pages, of which all the elements are searchable; articles, advertising and announcements. The supplements and FT Magazine are also included.

It will also soon be possible to cross-search the *Financial Times*, *The Times* and *The Economist*. The latter of which is available for the period 1843-2003, and soon up to 2006.

We were pleasantly surprised with the variety of search options offered by the database.

Obviously there is the basic search, which allows you to limit the search to Keywords/words in the title and sub-titles and authors, or you can search the entire text. You can also limit by date.

The advanced search offers many more possibilities.

You can combine multiple search lines separated by one of three Boolean options of your choice. Within the product, we have the option of using Boolean searches, proximity searches and truncation symbols (wildcards).

For each search line we can limit the search field if required.

We can also limit by date and type of information, but this last option didn't completely convince us.

On the other hand, an interesting option is being able to call on three levels of fuzzy logic – low, medium or high – to reflect the fact that OCR does not give reliable results 100% of the time for old documents.

In order to test this functionality, we entered the word *harbor* (port in American) in the title/sub-title, which gives 12 results.

The first level of fuzzy logic gives 1,890 results, mainly made up of the word *harbour* (port in English). The second level of fuzzy logic gave 5,877 results and the third 37,113. We do not know the algorithm used in the latter case, but it clearly has a very broad scan.

The results are shown with title, date, number of words and the document category. Clicking the title brings up the article, occasionally with the search term highlighted in green. As the text is not always easy to read with old documents, you can easily enlarge it.

You can, when viewing an article, see the titles of the other articles within the same edition: you can also view the official article citation, which can be saved or you can bookmark the web-address.

Universities and public institutions can

purchase the full archive, otherwise the cost of an annual subscription runs into hundreds of Euros, depending on the type of institution, unfortunately there is no other option but purchase or annual subscription.

This new database illustrates what we are finding more and more frequently with the old archives of the major newspapers from different countries.

In France, the online archives of *Le Monde*, for example, only go back to 1987 and those of *La Tribune* and the *Echos* to the beginning of the 90s.

Nonetheless there are within Gallica – the digital library of the BnF (Bibliothèque nationale de France) see Bases no. 256 – the archives of thirty titles from the national dailies, between the mid-nineteenth and mid-twentieth century.

---

**Further information**  
[www.gale.cengage.co.uk](http://www.gale.cengage.co.uk)